

Editorial

Algorithms for Managing, Querying and Processing Big Data in Cloud Environments

Alfredo Cuzzocrea

DIA Department, University of Trieste and ICAR-CNR, Trieste 34127, Italy; alfredo.cuzzocrea@dia.units.it;
Tel.: +39-040-558-3419; Fax: +39-040-558-3580

Academic Editor: Henning Fernau

Received: 12 January 2016; Accepted: 27 January 2016; Published: 1 February 2016

Big data (e.g., [1–3]) has become one of the most challenging research topics in current years. Big data is everywhere, from social networks to web advertisements, from sensor and stream systems to bio-informatics, from graph management tools to smart cities, and so forth. Cloud computing environments (e.g., [4–6]) represent the “natural” context for such data, as they embed several emerging trends, both at the research level and the technological level, which comprise high-performance, high reliability, high availability, transparency, abstraction, virtualization, and so forth.

At the convergence of these emerging trends, managing, querying and processing big data in Cloud environments, which have received a great deal of attention from the research community recently (e.g., [7–9]), plays a leading role, and algorithmic approaches to these challenges are very promising now. These approaches come from a rich variety of multi-disciplinary areas, ranging from mathematical models to approximation models, from resource-constrained paradigms to memory-bounded methods, and so forth. Furthermore, algorithms for managing big data according to a “systematic” view of the problem are gaining momentum. For instance, algorithms for efficiently managing MapReduce tasks over Clouds are a clear instance of the latter scientific area.

Inspired by these exciting research challenges, this special issue “Algorithms for Managing, Querying and Processing Big Data in Cloud Environments” of *Algorithms* focuses the attention on topics related to the theory and practice of algorithms for managing big data in Cloud environments, the design and analysis of algorithms for managing big data in Cloud environments, the tuning and experimental evaluation of algorithms for managing big data in Cloud environments, and so forth. The aim is that of providing a significant milestone on the road of the investigated topic, to be significant for both theory and practice, as well as applications and systems that are founded on such algorithms.

The special issue contains four papers which have been accepted after two rigorous review rounds. In the following, we provide an overview on these papers.

The first paper [10], entitled “Multiobjective Cloud Particle Optimization Algorithm Based on Decomposition”, by Li *et al.*, investigates the relevant multi-objective evolutionary paradigm based on decomposition (MOEA/D) that, as the authors correctly state, has received attention from many researchers in recent years. The paper thus presents a novel multi-objective algorithm based on decomposition and the Cloud computing model called multi-objective decomposition evolutionary algorithm based on Cloud Particle Differential Evolution (MOEA/D-CPDE). In the proposed method, the best solution found so far acts as a seed in each generation and evolves two individuals by a cloud generator. A new individual is produced by updating the current individual with the position vector difference of these two individuals. The performance of the proposed algorithm is verified on 16 well-known multi-objective problems, and experimental results indicate that MOEA/D-CPDE is competitive.

The second paper [11], entitled “Implementation of a Parallel Algorithm Based on a Spark Cloud Computing Platform”, by Wang *et al.*, proposes to parallelize the well-known MAX-MIN Ant System

(MMAS) algorithm in order to solve the annoying Traveling Salesman Problem (TSP) based on a Spark Cloud computing platform. Indeed, as authors correctly highlight, parallel algorithms, such as the ant colony algorithm, take a long time when solving large-scale problems. In the solution proposed by authors, MMAS is combined with Spark MapReduce to execute the path building and the pheromone operation in a distributed computer Cluster. In addition to this, to improve the precision of the proposed solution, the local optimization strategy 2-opt is adapted in MMAS. Experimental results show that Spark has a very great accelerating effect on the ant colony algorithm when the city scale of TSP or the number of ants is relatively large.

The third paper [12], entitled “A Data Analytic Algorithm for Managing, Querying, and Processing Uncertain Big Data in Cloud Environments”, by Jiang *et al.*, considers the problem of mining big data for supporting the discovery of useful information and knowledge. In this context, they propose a data analytic algorithm for managing, querying and processing transactions of uncertain big data in Cloud environments. The proposed framework, based on this algorithm, allows users to query these big data by specifying constraints expressing their interests, and processes the user-specified constraints to discover useful information and knowledge. Due to the fact that each item in every transaction in these uncertain big data is associated with an existential probability value expressing the likelihood of that item to be present in a particular transaction, computation could be intensive. In order to cope with this issue, the proposed algorithm makes use of the MapReduce model in a Cloud environment for effective data analytics on uncertain big data. Experimental results show the effectiveness of the overall solution.

Finally, the fourth paper [13], entitled “An Effective and Efficient MapReduce Algorithm for Computing BFS-based Traversals of Large-Scale RDF Graphs”, by Cuzzocrea *et al.*, focuses its attention on Resource Description Framework (RDF) graphs in terms of a relevant case of Big Web Data occurring in the so-called Semantic Web, leading to the well-known large-scale RDF graphs. They study the problem of effectively and efficiently computing traversals of large-scale RDF graphs over MapReduce and propose a solution that is based on the Breadth First Search (BFS) strategy for visiting (RDF) graphs to be decomposed and processed according to the MapReduce framework. The authors demonstrate how such implementation speeds up the analysis of RDF graphs with respect to competitor approaches. Experimental results clearly support the reliability of the provided contributions.

The described contributions still open the door for future research challenges to be investigated in order to further improve the management of big data in Cloud environments, yet they are inspired by previous research experiences in related scientific areas. For instance, big data compression (e.g., [14–17]) seems a promising solution to this end, as compressing data improves data management efficiency, but it has yet to be provided under well-defined probabilistic guarantees on deriving accuracy (e.g., [18–20]). Similarly, data fragmentation/partition techniques (e.g., [21–23]) should be considered as well, in the area of further solutions for improving performance while taking advantage of the typical distributed nature of Cloud platforms that, in our opinion, still expose interesting features not yet completely exploited beyond those of previous (distributed) settings (e.g., Grids, Clusters, and so forth).

In conclusion, there is still a lot of work to do in the context of managing, querying and processing big data in Cloud environments. We firmly hope this special issue represents a reliable milestone towards this difficult, yet exciting, research direction.

Acknowledgments: The Special Issue editor would like to express his gratitude to all contributors and reviewers whose efforts allowed making this special issue a success, as well as to the editorial staff of MDPI Algorithms Journal for their continuous and diligent assistance whenever needed.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Agrawal, D.; Das, S.; El Abbadi, A. Big Data and Cloud Computing: Current State and Future Opportunities. Available online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.188.5819> (accessed on 29 January 2016).
2. Chen, C.L.P.; Zhang, C.-Y. Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. *Inf. Sci.* **2014**, *275*, 314–347. [[CrossRef](#)]
3. Cuzzocrea, A.; Saccà, D.; Ullman, J.D. Big Data: A Research Agenda. In Proceedings of the 17th International Database Engineering & Applications Symposium (IDEAS'13), Barcelona, Spain, 9–11 October 2013; pp. 198–203.
4. Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A.D.; Katz, R.; Konwinski, A.; Lee, G.; Patterson, D.; Rabkin, A.; Stoica, I.; et al. A View of Cloud Computing. *Commun. ACM* **2010**, *53*, 50–58. [[CrossRef](#)]
5. Buyya, R.; Yeo, C.S.; Venugopal, S.; Broberg, J.; Brandic, I. Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. *Future Gener. Comput. Syst.* **2009**, *25*, 599–616. [[CrossRef](#)]
6. Cuzzocrea, A. Analytics over Big Data: Exploring the Convergence of Data Warehousing, OLAP and Data-Intensive Cloud Infrastructures. In Proceedings of the 2013 IEEE 37th Annual Computer Software and Applications Conference (COMPSAC'13), Kyoto, Japan, 22–26 July 2013; pp. 481–483.
7. Baker, T.; Al-Dawsari, B.; Tawfik, H.; Reid, D.; Ngoko, Y. GreeDi: An Energy Efficient Routing Algorithm for Big Data on Cloud. *Ad Hoc Netw.* **2015**, *35*, 83–96. [[CrossRef](#)]
8. Zhang, L.; Li, Z.; Wu, C.; Chen, M. Online Algorithms for Uploading Deferrable Big Data to the Cloud. In Proceedings of the IEEE 2014 International Conference on Computer Communications (INFOCOM'14), Toronto, ON, Canada, 27 April 2014–2 May 2014; pp. 2022–2030.
9. Yu, B.; Cuzzocrea, A.; Jeong, D.H.; Maydehura, S. A Bigtable/MapReduce-Based Cloud Infrastructure for Effectively and Efficiently Managing Large-Scale Sensor Networks. In Proceedings of the 5th International Conference on Data Management in Cloud, Grid and P2P Systems (GLOBE'12), Vienna, Austria, 5–6 September 2012; pp. 25–36.
10. Li, W.; Wang, L.; Jiang, Q.; Hei, X.; Wang, B. Multiobjective Cloud Particle Optimization Algorithm Based on Decomposition. *Algorithms* **2015**, *8*, 157–176. [[CrossRef](#)]
11. Wang, L.; Wang, Y.; Xie, Y. Implementation of a Parallel Algorithm Based on a Spark Cloud Computing Platform. *Algorithms* **2015**, *8*, 407–414. [[CrossRef](#)]
12. Jiang, F.; Leung, C.K. A Data Analytic Algorithm for Managing, Querying, and Processing Uncertain Big Data in Cloud Environments. *Algorithms* **2015**, *8*, 1175–1194. [[CrossRef](#)]
13. Cuzzocrea, A.; Cosulschi, M.; De Virgilio, R. An Effective and Efficient MapReduce Algorithm for Computing BFS-Based Traversals of Large-Scale RDF Graphs. *Algorithms* **2016**, *9*, 7. [[CrossRef](#)]
14. Yang, C.; Zhang, X.; Zhong, C.; Liu, C.; Pei, J.; Ramamohanarao, K.; Chen, J. A Spatiotemporal Compression Based Approach for Efficient Big Data Processing on Cloud. *J. Comput. Syst. Sci.* **2014**, *80*, 1563–1583. [[CrossRef](#)]
15. Cuzzocrea, A.; Furfaro, F.; Masciari, E.; Saccà, D.; Sirangelo, C. Approximate Query Answering on Sensor Network Data Streams. In *GeoSensor Networks*; Stefanidis, A., Nittel, S., Eds.; CRC Press: Cleveland OH, 2004; pp. 53–72.
16. Cuzzocrea, A. Accuracy Control in Compressed Multidimensional Data Cubes for Quality of Answer-based OLAP Tools. In Proceedings of the 18th International Conference on Scientific and Statistical Database Management (SSDBM'06), Vienna, Austria, 3–5 July 2006; pp. 301–310.
17. Cuzzocrea, A. Improving Range-Sum Query Evaluation on Data Cubes via Polynomial Approximation. *Data Knowl. Eng.* **2006**, *56*, 85–121. [[CrossRef](#)]
18. Xi, R.; Lin, N.; Chen, Y. Compression and Aggregation for Logistic Regression Analysis in Data Cubes. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 479–492.
19. Cuzzocrea, A.; Furfaro, F.; Saccà, D. Hand-OLAP: A System for Delivering OLAP Services on Handheld Devices. In Proceedings of the 6th IEEE International Symposium on Autonomous Decentralized Systems (ISADS'03), Pisa, Italy, 9–11 April 2003; pp. 80–87.

20. Cuzzocrea, A.; Matrangolo, U. Analytical Synopses for Approximate Query Answering in OLAP Environments. In Proceedings of the 15th International Conference on Database and Expert Systems Applications (DEXA'04), Zaragoza, Spain, 30 August 2004–3 September 2004; pp. 359–370.
21. Daenen, J.; Neven, F.; Tan, T. Gumbo: Guarded Fragment Queries over Big Data. In Proceedings of the 18th International Conference on Extending Database Technology (EDBT'15), Brussels, Belgium, 23–27 March 2015; pp. 521–524.
22. Cuzzocrea, A.; Saccà, D.; Serafino, P. A Hierarchy-Driven Compression Technique for Advanced OLAP Visualization of Multidimensional Data Cubes. In Proceedings of the 8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'06), Krakow, Poland, 4–8 September 2006; pp. 106–119.
23. Cuzzocrea, A.; Darmont, J.; Mahboubi, H. Fragmenting Very Large XML Data Warehouses via K-Means Clustering Algorithm. *Int. J. Bus. Intell. Data Min.* **2009**, *4*, 301–328. [[CrossRef](#)]



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).